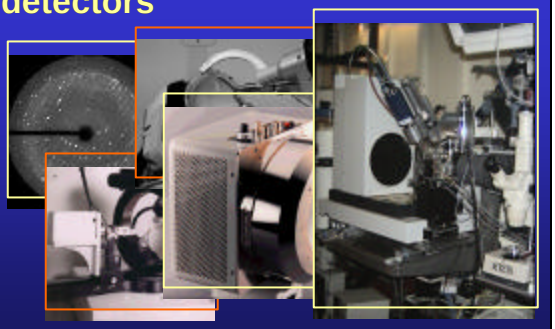
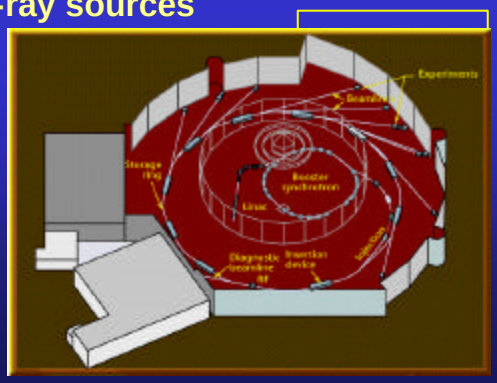
Crystallography borders, naturally, on pure physics, chemistry, biology, mineralogy, technology and also on mathematics, but is distinguished by being concerned with the methods and results of investigating the arrangement of atoms in matter, particularly when that arrangement has regular features.

Paul P. Ewald, Acta Crystallographica (1948), ***1****, 2.*

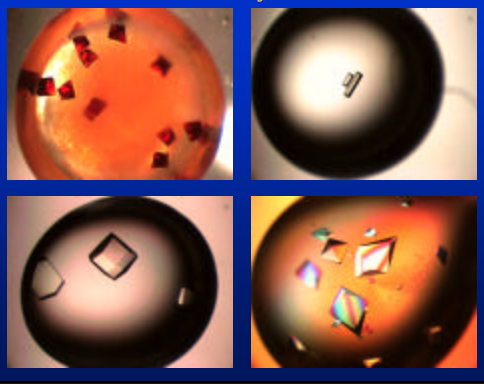
detectors



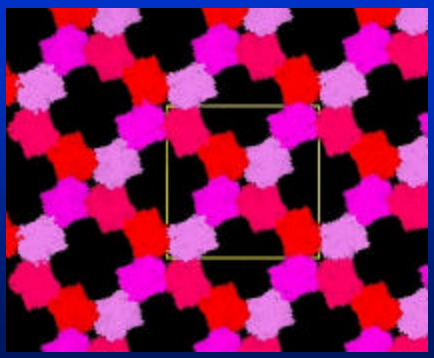
x-ray sources



What are crystals ? Kirsten Böttcher



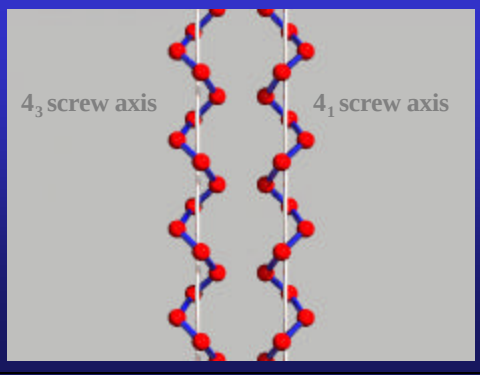
What are crystals ? Ralph Krätzner



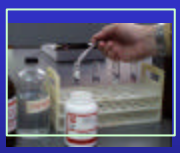
Crystal symmetry

- There are exactly 230 different ways of combining symmetry elements in space (the space groups) so that the same pattern is repeated for ever in all directions.
- Of these 230, only 65 are chiral and so are suitable for proteins, DNA, RNA, sugars etc.
- The symmetry elements in the chiral space groups are either 2-, 3-, 4- or 6-fold rotations about an axis or simultaneous rotation and translation along the axis (giving an infinite spiral, e.g. of molecules linked by H-bonds).

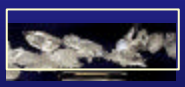
Crystal symmetry



crystallization and sample mounting

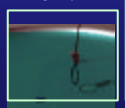


Crystals are produced by recrystallization from specific solvent(s).



Single crystals must be selected and mounted on glass fibers.

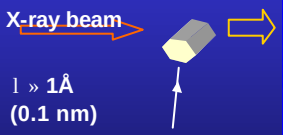
Linbro plate and microbridges hold sitting drops.



Crystals form in the drops by vapor diffusion.

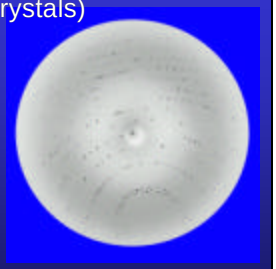
Crystals are removed from drops and transferred into cryoprotectants using cryoloops.

X-ray diffraction
(protein crystals)



$\lambda \approx$ **1Å (0.1 nm)**

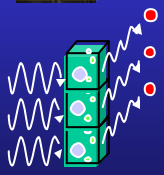
~ (0.2mm)^3 crystal ~10^{13} unit cells, each ~ (100Å)^3



Diffraction pattern on CCD or image plate

What determines the *intensity* of a reflection ?

The intensity of each reflection results from simple super-position of the individual atomic scattering contributions.



$$F_{hkl} = \sum_{i=1}^{n} f^{\,i} \cdot e^{2pi(hx_i + ky_i + lz_i)}$$

Electronic property of atom

two terms

Structural property (position)

Diffraction by atoms and molecules

1 atom

FT

FT

Fourier transform of 1 atom

1 mol.

FT

FT

Fourier transform of one molecule

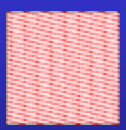
Kevin Cowtan www.ysbl.ac.uk/~cowtan

Diffraction by lattices and crystals

lattice

FT

FT



Fourier transform of lattice

crystal

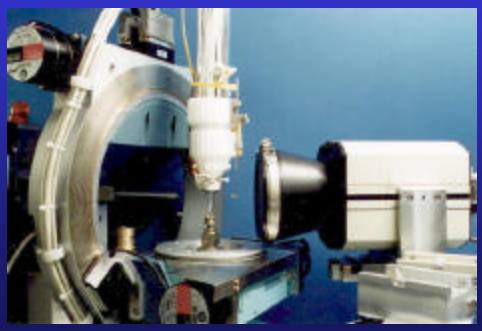
FT

FT

Fourier transform of crystal

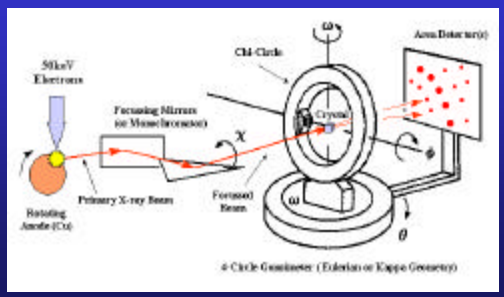
Kevin Cowtan www.ysbl.ac.uk/~cowtan/

The diffraction experiment



experimental set-up

(Crystallography 101 http://www-structure.llnl.gov)



Bragg's law

$$n\lambda = 2d\sin(\theta)$$

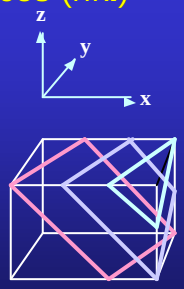
We can think of diffraction as reflection at sets of planes running through the crystal. Only at certain angles 2θ are the waves diffracted from different planes a whole number of wavelengths apart, i.e. in phase. At other angles the waves reflected from different planes are out of phase and cancel one another out.

Reflection indices (hkl)

These planes must intersect the cell edges rationally, otherwise the diffraction from the different unit cells would interfere destructively.

We can index them by the number of times *h, k and* ℓ that they cut each edge.

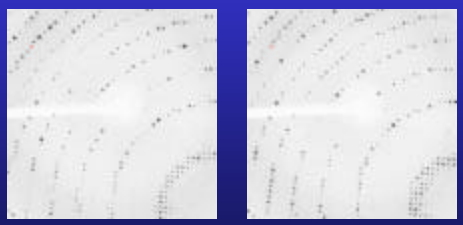
The same *h, k and* ℓ values are used to index the X-ray *reflections* from the planes.



Planes 3 -1 2 (or -3 1 -2)

Diffraction patterns

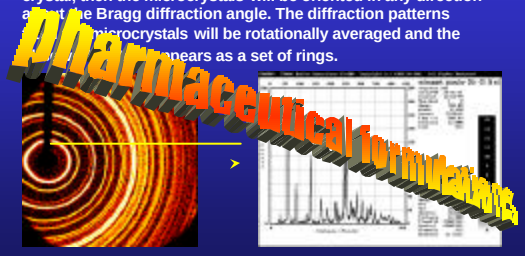
Two successive CCD detector images with a crystal rotation of one degree per image



For each X-ray reflection (black dot) indices *h,k,* ℓ can be assigned and an intensity $I = F^2$ measured

Diffraction patterns

If the crystalline sample is microcrystalline and not a single crystal, then the microcrystals will be oriented in any direction about the Bragg diffraction angle. The diffraction patterns from microcrystals will be rotationally averaged and the pattern appears as a set of rings.



For each ring, we can integrate and assign an intensity *I* as a function of 2θ .

Reciprocal space

- The immediate result of the X-ray diffraction experiment is a list of X-ray reflections *hkℓ* and their intensities *I*.
- We can arrange the reflections on a 3D-grid based on their *h*, *k* and *ℓ* values. The smallest repeat unit of this *reciprocal lattice* is known as the *reciprocal unit cell*; the lengths of the edges of this cell are inversely related to the dimensions of the real-space unit cell.
- This concept is known as *reciprocal space*; it emphasizes the inverse relationship between the diffracted intensities and *real space*.

The *resolution* of the X-ray data

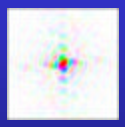
If we apply Bragg's law to the maximum 2θ to which reflections were measured, we obtain an expression for the maximum resolution (smallest) *d*:

$$d = \lambda / 2\sin(\theta_{max})$$

As it happens, *d* corresponds quite well to the minimum distance apart at which two equal atoms can be resolved in an electron density map.

A low resolution duck

Fourier transform of duck



FT

FT

duck

FT of duck with high resolution data removed

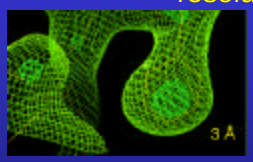
FT

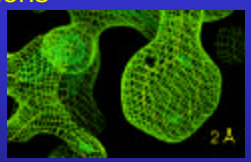
FT

duck at low resl.

Kevin Cowtan www.ysbl.ac.uk/~cowtan/

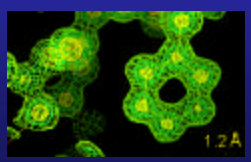
The same map at different resolutions





At 2Å one can often identify side-chains.

At 1.2Å bonded atoms give separate peaks.



Bernhard Rupp http://www-structure.llnl.gov/

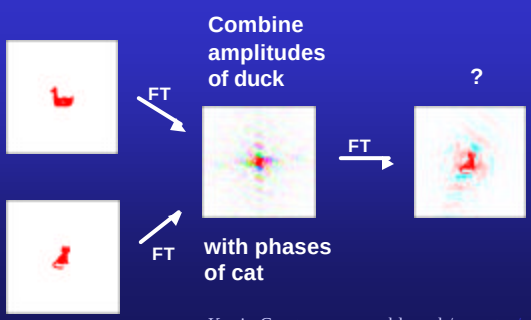
The structure factor *F* and electron density *r*

$$F_{hkl} = \int_V r_{xyz} \exp[+2\pi i(hx+ky+\ell z)]\, dV$$

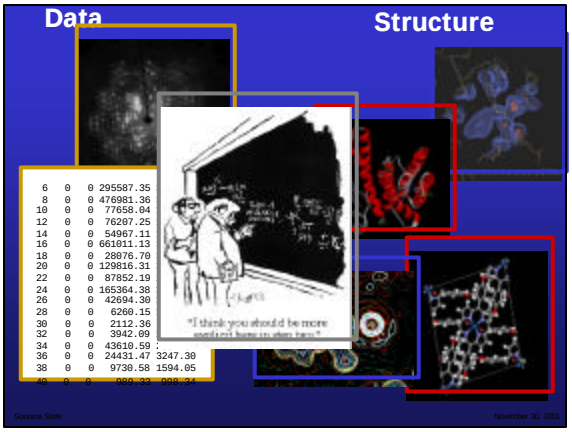
$$r_{xyz} = (1/V) \sum_{hkl} F_{hkl} \exp[-2\pi i(hx+ky+\ell z)]$$

F **and** r **are inversely related by these Fourier transformations. Note that** r **is real and positive but** ***F*** **is a complex number: in order to calculate the electron density from the diffracted intensities** $I = F^2$ **we need the PHASE (**f**) of** ***F*****. Unfortunately it is almost impossible to measure** f **directly!**

Amplitudes (F) and phases (f)



Kevin Cowtan www.ysbl.ac.uk/~cowtan



The crystallographic phase problem

- In order to calculate an electron density map, we require both the intensities $I = F^2$ and the phases f of the reflections $hk\ell$.
- The information content of the phases is appreciably greater than that of the intensities.
- Unfortunately, it is almost impossible to measure the phases experimentally !

This is known as the *crystallographic phase problem* and would appear to be insoluble.

Solving the phase problem

- Structures of <1000 atoms can be solved by brute force statistical *Direct Methods*, given data to 1.2Å or better. Most small molecule structures are solved this way.
- A closely related known structure can be used as a search fragment for MR = *Molecular Replacement*.
- Heavy atoms can be introduced and the small changes in the reflection intensities exploited (SIR / MIR = Single / Multiple *Isomorphous Replacement*).
- Heavier atoms exhibit wavelength-dependent *Anomalous Scattering*, with the result that F_{hkl} and F_{-h-k-l} are not exactly equal. These small differences can be exploited in the SAD and MAD methods. For MAD, either metal atoms such as Fe present in the protein, or Se in selenomethionine (genetically modified methionine) are suitable anomalous scatterers.

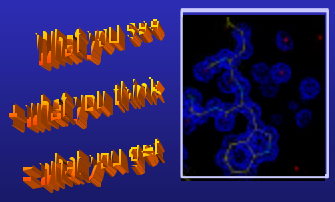
Molecular Replacement (MR)

- Requires a search fragment that is fairly similar to the structure being investigated (at least about 30% sequence identity). A related crystal structure is usually much more effective than an NMR or theoretically modeled structure.
- Does not require high resolution data.
- Very useful for complexes and mutants.
- Gives atom positions ready for refinement without the need to trace the chain etc.
- BUT can suffer seriously from MODEL BIAS !

'Heavy atom' methods

- Soaking the crystals with heavy atom salts (Pt, Pb, Hg, U etc. - Hg attaches to free cysteines) is the classical approach but deterioration of the crystals and lack of isomorphism are major problems.
- Higher quality maps can be obtained with MAD (multiple wavelength anomalous dispersion) phasing for which all the data can be collected from one frozen crystal by varying the synchrotron wavelength. The MAD data can be analyzed to extract the heavy atom structure factors and the phase shifts between the heavy atom and protein phases, so the heavy atoms (which have to be found by e.g. direct methods) provide reference phases.
- SAD (single wavelength anomalous diffraction) followed by *density modification* can also give interpretable maps in many cases, even without a synchrotron.

Once initial phases are obtained, a model, which makes chemical and physical sense (your perception of what is right), is fitted into the initial electron density map (the actual and sometimes marginal experimental reality)



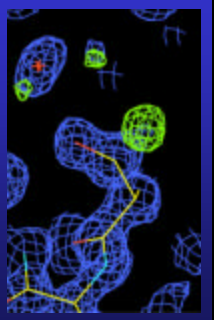
Structure refinement

The model is usually improved by least-squares minimization of:

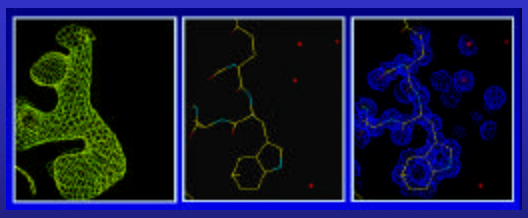
$$M = \sum_{hkl} w(F_o - F_c)^2 + \sum_t w(Y_t - Y)^2$$

alternately with rebuilding of the model by interactive graphics.

w are weights, F_o and F_c are the observed structure factors (without phase !) and Y and Y_t are the current and target values of *restraints* such as bond lengths and angles. The lower the resolution is, the more important are these restraints.



Model bias ?



Measured F's combined with expt. phases | **Model (hypothetical)** | **Measured F's phased by model**

Bernhard Rupp www-structure.llnl.gov/

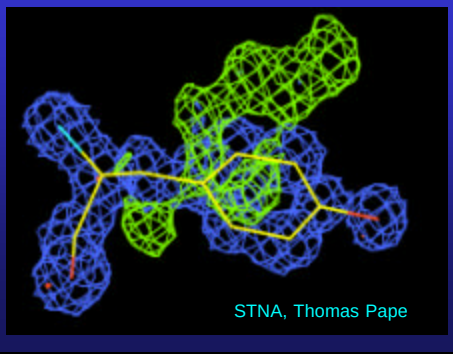
Figures of merit

The quality of a crystal structure refinement used to be measured with the index $R = \sum_{hkl} ||F_o|-|F_c|| / \sum_{hkl} |F_o|$

However *R* can be reduced artificially by refining more parameters, so now it is usual to reserve 5 to 10% of the reflections to calculate an index R_{free} (same formula). Since the model is not refined against these reflections they provide an independent assessment of the quality of the structure.

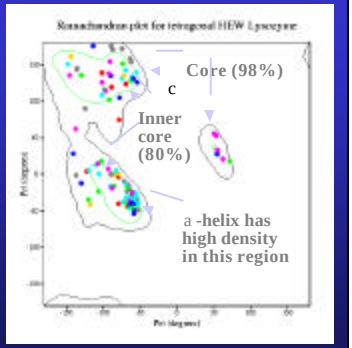
The refinement is completed when it is not possible to reduce R_{free} further and there is no significant remaining difference electron density or disagreements with the restraint target values. Usually one is happy if *R* is below 20% and R_{free} is below 25%. General rule: R in % should be about 10 x the highest resolution.

Multiple conformations



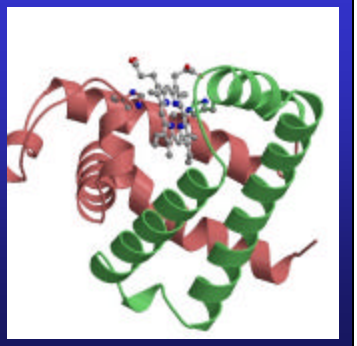
STNA, Thomas Pape

Since the torsion angles phi and psi are not restrained in the refinement, the Ramachandran *psi/phi* plot is a good diagnostic. 98% of points should fall in the core region, 80% in the inner core.



Myoglobin

The oxygen transport protein Myoglobin was the first protein structure to be determined. The 2Å structure by Kendrew et al. (1960) was one of the foundations of molecular biology.



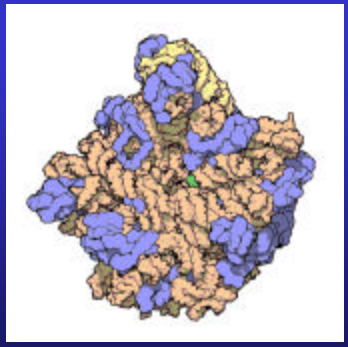
Crystal structures of the ribosome

The ribosome consists of two subunits, which lock together with a groove between them along which runs the messenger RNA. The ribosome walks along the tRNA 3 nucleotides in each step, building a new protein one residue per step.

The large subunit is composed of two RNA strands and some proteins (colored blue) that stabilize the structure.
Structures of ribosome complexes strongly indicate that RNA, not protein, is the active catalyst, so the ribosome is a ribozyme ! This supports the theory that RNA came before proteins in the course of evolution.

The large ribosome subunit

This ca. 100000 atom structure was determined (in 2000) to a resolution of 2.4Å !



http://www.rcsb.org

Summary

Most of the beautiful schematic pictures of proteins in textbooks of chemistry and molecular biology represent structures determined by X-ray diffraction. But do not be taken in by the artistic effects, not all crystal structures are equally precise or reliable ! The resolution of the X-ray data, the agreement of the model with the data (R_{free}), the validation of the structure using independent knowledge (Ramachandran) and the possibility of model bias should be considered too !